DISSECTING THE ROLE OF NANOG IN STEM CELL PLURIPOTENCY *IN SILICO*

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENT FOR THE DEGREE OF

Bachelor of Technology

in

BIOTECHNOLOGY

by

SANJIBITA MISHRA

110BT0010

Under the Supervision of

Dr. MUKESH KUMAR GUPTA

Department of Biotechnology and Medical Engineering

National Institute of Technology Rourkela

Rourkela, Odisha, 769 008, India

May 2014

Certificate

This is to certify that the thesis entitled "**DISSECTING THE ROLE OF NANOG IN THE STEM CELL PLURIPOTENCY *IN SILICO***" by **SANJIBITA MISHRA(110BT0010),** in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in BIOTECHNOLOGY during session 2010-2014 in the Department of Biotechnology and Medical Engineering, National Institute of Technology Rourkela, is an authentic work carried out by her under my supervision and guidance. To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/Institute for the award of any degree or diploma.

Place: NIT Rourkela
Date: 12th May 2014

Dr. M. K Gupta
Associate Professor
Biotechnology and Medical Engineering
National Institute of Technology
Rourkela-769 008, Odisha (India)

Acknowledgement

Successful completion of this project is the outcome of guidance and assistance from Faculty and friends and I am extremely fortunate to have got them in helping me complete my project.

I avail this golden opportunity to extend my sincere gratitude to my guide, **Prof. Mukesh kumar Gupta**, Department of Biotechnology and Medical Engineering, NIT Rourkela for his invaluable academic and professional guidance, encouragement and motivation at each stage of this project. I put forth my sincere gratitude to **Prof. Krishna Pramanik**, Head of Department,

Department of Biotechnology and Medical Engineering, NIT Rourkela for her constant encouragement and valuable departmental facilities.

.I would like to thank Ms. JyotiBala, M.Tech Scholar, Department of Biotechnology and Medical Engineering, NIT Rourkela for their regular support, help and motivation. I would also thank my Alma matter and faculty members without whom this project would have been seen completion. I also extend my thanks to my family and friends.

Finally I would like to express my heartiest thank to my friends Pranjali Nanda, Satyajit Das , DuttaSugave and AniketMazumdar for their constant help during the project. I thank them for being with me throughout the completion of project during hours of despair and dismay.

Place: NIT Rourkela

Sanjibita Mishra

Date: 12th May 2014

110BT0010

Biotechnology and Medical Engineering

National Institute of Technology

Rourkela-769 008, Odisha (India)

Contents

	Page No.
➢ Certificate	1
➢ Acknowledgement	2
➢ List of Tables	5
➢ List of Figures	5
➢ Abstract	6
➢ **Chapter 1** – Introduction	7-10
➢ **Chapter 2** – Literature Review	11-13
➢ **Chapter 3** – Materials and methods	14-24
3.1 OVERVIEW	14-18
3.2 TUXEDO PROTOCOL	19-21
3.3 R	22
3.4 DAVID	23
3.5 CYTOSCAPE	24
Chapter 4 –Results and Discussion	25-40
➢ Conclusion	41
➢ References	42-43

List of Tables

SL NO.	TABLE NO.	TABLE CAPTION	PAGE NO.
1.	I	Table of Data Sample Used	**16**
2.	II	Table for Comparative interaction between different conditions	**25**
3.	III	Table for differentially expressed gene	**26**

List of Figures

SL. NO.	FIG. NO.	FIGURE CAPTION	PAGE NO.
1.	Fig 1.	RNA-Seq Flow cell	**17**
2.	Fig 2.	Tuxedo Protocol	**19**
3.	Fig 3.	Cytoscape Gene Cluster	**33**

ABSTRACT:

Nanog is associated with the pluripotency and constitute an important factor in cellular reprogramming. Thus, dissecting the molecular function of Nanog would provide better understanding of the process of reprogramming, cellular replication and differentiation. This study was designed to utilize the Next Generation Sequencing (RNA-Seq) data to identify, discover and profile the RNA transcripts influenced by RNA silencing of Nanog in mouse embryonic stem (ES) cell. Data from mouse induced pluripotent stem (iPS) and mouse embryonic fibroblast (MEF) cells were used as controls. GALAXY, an online open source software platform, was used and Tuxedo Protocol was followed to identify the differentially expressed genes and transcripts. The differentially expressed genes and transcripts were functionally annotated for biological and molecular functions and identify splice junction, exon-intron boundaries, and identification of primary transcripts. Results showed that, in Nanog null cell lines, Sgol2, Mcm7, Psma4 were up-regulated by a factor of 1.36 whereas several pluripotency related genes were down regulated. The up-regulated genes were associated with the regulation of cell cycle and regulation of pre replication complex whereas down-regulated genes were found to be associated with protein synthesis pathways, cellular proliferation and differentiation pathways besides pluripotency. Further, upon Nanog knock-down, genes associated with the process of replication were activated while genes associated with the cellular differentiation were down-regulated by a relative expression in the range of (0.3-0.9). In conclusion, the results of the present study suggests that Nanog may not only be responsible for determining the pluripotency and reprogramming of cells but also for the process of cellular replication .

Keywords: Cellular Reprogramming, Differential expression, Pluripotent cells, Cell differentiation ,Cell replication.

CHAPTER ONE: INTRODUCTION

INTRODUCTION:

RNA –Seq is also known as whole transcriptome shotgun sequencing , is a method which uses the advancement in the field of next generation sequencing to reveal the amount of RNA transcripts and quality of genome at a particular moment in time. It is a powerful technology to identify, discover,profile and quantify RNA transcripts. RNA-Seq provides the most precise measurement of levels of transcriptome and their isoforms. And is used for transcriotome profiling using deep sequencing

A transcriptome represents that small percentage of the genetic code that is transcribed into RNA molecules — estimated to be less than 5% of the genome in humans (Frith *et al.*, 2005). Transcriptome is the complete set of transcripts in a cell and their quantity , for a specific developmental phase or physiological condition. Understanding the transcriptome , interpreting transcripts is crucial to understand the molecular function of cellular constituents and to gain a deeper knowledge of diseases and their cause. The main aim of transcriptome profiling is to identify, quantify and characterise all types of RNA present in cell including mRNAs, non-coding RNA, rRNA, tRNA . RNA-Seq can also be used to determine exon and intron boundaries and verify or amend previously annotated 5' and 3' gene boundaries. RNA-Seq remains the standard method for identifying and annotating coding and non-coding genes alike.RNA-Seq research is of prime importance in identifying cellular pathway alterations and gene expression level changes in cancer studies.

Prior to advent of Next generation sequencing transcriptome profiling was carried out with expression microarrays and serial analysis of gene expression(SAGE).Multiple technologies have been developed to identify and quantify the transcriptome, including hybridization-or sequence-based approaches. Hybridization-based approaches involve incubating fluorescently labelled cDNA with custom-made microarrays and commercial available oligo

microarrays. Microarrays have been designed as arrays with probes to identify exon junctions and to quantify spliced isoforms. Genomic tiling microarrays represent the genome at high density and allow the mapping of transcribed regions at a very high resolution. Hybridization-based approaches are relatively inexpensive except for transcriptome tilling of large genome. These methods have limitations, such as : dependence on existing genome sequence knowledge; high background noise level, restricted dynamic range of detection due to background and saturation of signals, difficulty in comparing expression levels across different experiments due to varying normalization methods[1].

Initially Sanger based EST libraries were used but these techniques were expensive ,had less throughput and were not quantitative .To overcome these limitations Tag based methods such as Cap analysis of gene expression(CAGE) ,Serial analysis of gene expression(SAGE) and Massively parallel signature sequencing(MPSS) analysis were developed. Tag based methods were able to provide precise gene expression level, but were not preferred as were dependent on expensive SANGER sequencing and due to non distinguishable limits of isoform detection. These short comings limited the use of traditional sequencing technology and paved way for the use of RNA-Seq to annotate transcriptomes.

In RNA-Seq the first step is the preparation of RNA Poly(A)library , RNA is converted to cDNA fragments using ligators and adaptors attached to both ends. Poly(T) magnetic beads are used to separate coding RNA from non coding RNA. The RNA is converted to cDNA by the process of reverse transcription. Each transcript, with or without amplification, is sequenced by high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). The resulting sequence reads are aligned with the reference genome or transcriptome, and are segregated into three types: exonic reads, junction reads and poly(A) end-reads. These three groups are used to generate a base-resolution expression profile for each gene. Commonly used technology for RNA-Seq are

Illumina IG, Applied Biosystems SOLiD and Roche 454 Life Sciencesystems . The transcriptome assembly of raw reads can be achieved by two methods :de-novo and genome guided. Organism for which the genome is unavailable or highly fragmented ,de novo assembly methods are used to construct reference transcriptome , based on similarity in local regions.The detection by genome guided method is based on algorithms for Fischer exact test, Likelihood ratio test, edger, DESeq, BsySeq, Two stage Poisson model[2]. Transcript Representation and enumeration can be done by Overlap graph , Connectivity graph ,Splice graph and Sub-exon graph .Transcript selection is achieved via the process of numerical selection using Quadratic programming, Expectation Maximisation.[3]

RNA-Seq has increased the plethora of knowledge in the field of transcriptomics,it is particularly attractive for non –model organisms with genomic sequences yet to be determined. RNA-Seq is helpful in detecting the precise location of transcription boundaries. It has very low,rather nil background signal because DNA regions can be unambiguously mapped to specific regions in the genome. RNA-Seq has large dynamic range and specificity. It shows high level of reproducibility with Biological and Technical replicates.

RNA-Seq has wide landscape of applications ranging from Annotation of novel genes, splicing events, exon intron boundaries, transcription boundaries, coding and non coding RNA, identifying Single Nucleotide Polymorphism (SNP), highly accurate and reliable RNA quantification, Gene differential expression .

CHAPTER TWO: LITERATURE REVIEW

LITERATURE REVIEW:

Nanog , a homeodomain-containing protein, is a transcription factor which is deterministic of Pluripotency of embryonic stem cell[4]. And is responsible for the self renewal property of undifferentiated Embryonic Stem Cells . Nanog over expressed activates cells to enter into S phase of cell cycle and promotes proliferation. It works in concert with oct-4 and SOX 2 to establish embryonic stem cell identity. Nanog also acts like transcriptional activator for Rex1. Nanog Knock down promotes the process of cellular differentiation ,thereby causing

the loss of pluripotency [5]. It has been experimentally shown that p53 binds to the promoter of Nanog and prevents its expression after DNA damage. Hence p53 acts antagonistic to the functioning of Nanog [10]. Nanog protein recognizes particular DNA Sequence which is found in the promoter region of Gata6.Gata6 plays a crucial role in primitive endoderm differentiation[9].Hence primary function of Nanog is to prevent the differentiation of Embryonic Stem Cell into endoderm[11].Nanog determines cell fate in both Embryonic stem cell and Cancer stem cell. Unregulated Nanog is associated with poor prognosis of various cancer. Nanog has been found to play crucial role in the networks which regulate tumour cell development, Epithelial-Mesenchymal transition, immuno suppressive , drug resistance ,proliferation, self-renewal, motility. These properties are characteristic of cancer stem cell. Nanog plays an important role in the conversion of fibroblast (Somatic cells) to Embryonic Stem Cells. Hence is responsible for the formation of induced pluripotent stem cells[12]. Although Nanog is necessary for cellular reprogramming ,use of canonical transcription factors like c-myc ,klf4 , Sox2 ,Oct4 under standard conditions does generate pluripotent cells from somatic cells[7]. Expression of Nanog has been detected in many germ cell tumours including ovarian, prostrate, brain, oral, kidney, cervix and breast cancer[13],[12],[15][16],[17].Over expression of Nanog has been associated with increased

resistance to drug in certain type of breast cancer[12]. Nanog is also over expressed in ovarian cancer cells having metastatic property. Knock down of Nanog has been associated with impediment of proliferation, migration, invasion (Siu MK et al.,2013).Nanog and oct4 are often considered indispensible for pluripotency. But it has been found that even in absence of Nanog mouse embryonic fibroblast can be transformed into cells having self renewable capability like Embryonic Stem Cells formation of induced pluripotent stem cells,hence it can be deduced that although Nanog is one of the governing factors to determine pluripotency ,it is not required for the generation of iPSC[7]. As Nanog has been identified as a therapeutic target for cancer and has been associated with wide spread use in regenerative medicine,it would be beneficial to identify the genes which have been involved in the induction of adult cells into pluripotent stem cells .As it would provide better perspective on the genes involved in the process of differentiation and the ones associated with networks responsible for maintaining pluripotency.

CHAPTER THREE: MATERIALS AND METHODS

MATERIALS AND METHODS:

The data used to understand the role of Nanog in stem cell pluripotency Insilco was retrieved from Gene Expression Omnibus (GEO) database from data entry –**GSE53212**. The data was released under the title "Nanog Independent Reprogramming to iPSCs with Canonical Factors". The data consisted of 5groups with (2-4)biological replicates.Platforms used were: Illumina HiSeq 2000(mouse) and Illumina HiSeq 2500(mouse).The murine used for experimentation were of genotype :WT and strain:129/B6.To obtain Nanog null Mouse Embryonic Fibroblast(MEF),Nanog null Embryonic Stem Cell (ESC) were injected into blastocyst stage embryos. For reprogramming of Mouse Embryonic Fibroblast(MEF) into Induced Pluripotent Stem Cell(iPSC), MEFs were transduced with retroviruses carrying murine Oct4, Sox2, Klf4, +/- c-Myc and were cultured in 2i+LIF conditions .MEF were grown in 10% FBS[14].RNA-Seq libraries were generated by using 250 ng of total RNA using the Illumina TruSeq RNA kit v2 and base calls were performed using CASAVA.

Total 6 comparisons done between the Nanog (+/+) cell lines and Nanog(-/-) cell lines.The comparisons were done to generate reference annotation transcript for further comparison between Nanog null cell lines and Nanog positive cell lines.

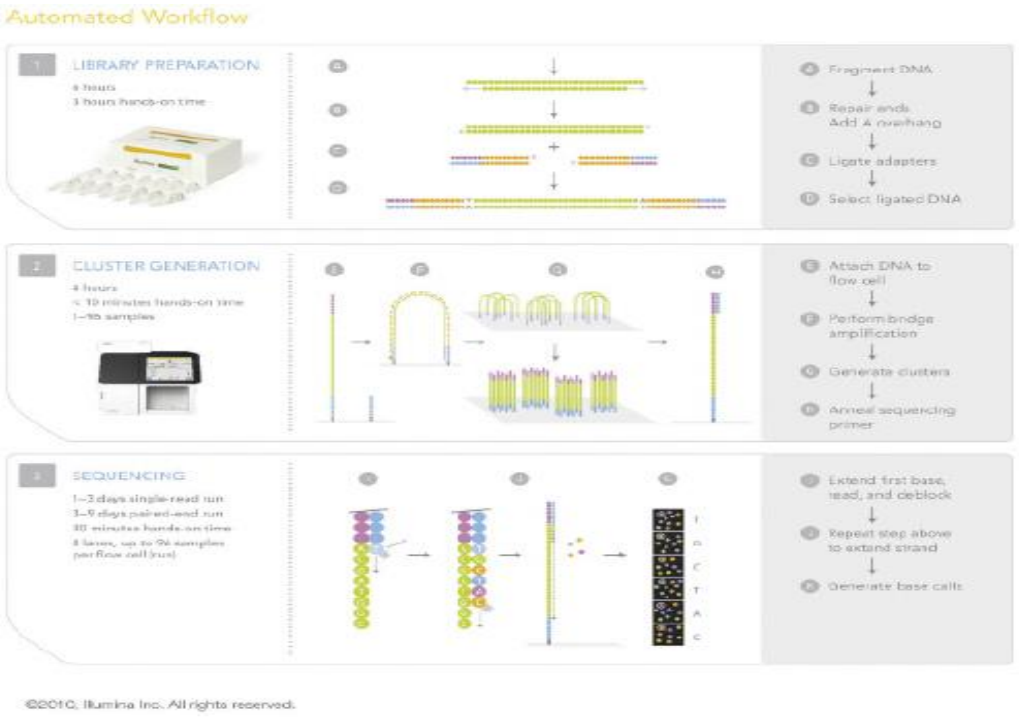
- Embryonic Stem Cell Vs Induced Pluripotent Stem Cell
- Embryonic Stem Cell Vs Mouse Embryonic fibroblast
- Mouse Embryonic fibroblast Vs Induced Pluripotent Stem Cell
- Nanog(-/-)Embryonic Stem Cell Vs Nanog(-/-)Induced Pluripotent Stem Cell
- Nanog(-/-)Embryonic Stem Cell Vs Embryonic Stem Cell
- Nanog(-/-)Embryonic Stem Cell Vs Induced Pluripotent Stem Cell

Sample Name	Cell Type	Genotype	Strain	Passages
ESC	Embryonic Stem Cell	WT	129/B6	10-15
iPSC	Induced pleuripotent Stem cell	WT	129/B6	6
MEF	Mouse Embryonic fibroblast	WT	129/B6	2
Nanog(-/-) ESC	Embryonic stem cell	WT	129/B6	10-15
Nanog(-/-) iPSC	Induced pleuripotent stem cell	WT	129/B6	6

Illumina Sequencing Technology provides highest data accuracy, simple workflow, and a broad range of applications. It leverages clonal array formation and reversible terminator technology for rapid, accurate large scale sequencing. The basic steps include:

- Cluster generation: cDNA templates are immobilised on the flow cell surface. Solid phase amplification is performed and 1,000 identical copies are created from each template transcript during each cycle.
- Sequencing Synthesis: uses four fluorescently labelled nucleotide Fluorescent dye is use to identify the base and in enzymatically cleaved to allow incorporation of next nucleotide. Base calls are made directly from signal intensity measurements.
- Analysis pipeline: Is characterised by deep sampling and uniform coverage. Deep Sampling uses weighted majority and statistical sampling. Each base is associated

with a quality score,to which the software applies weighting factor to generate confidence scores.



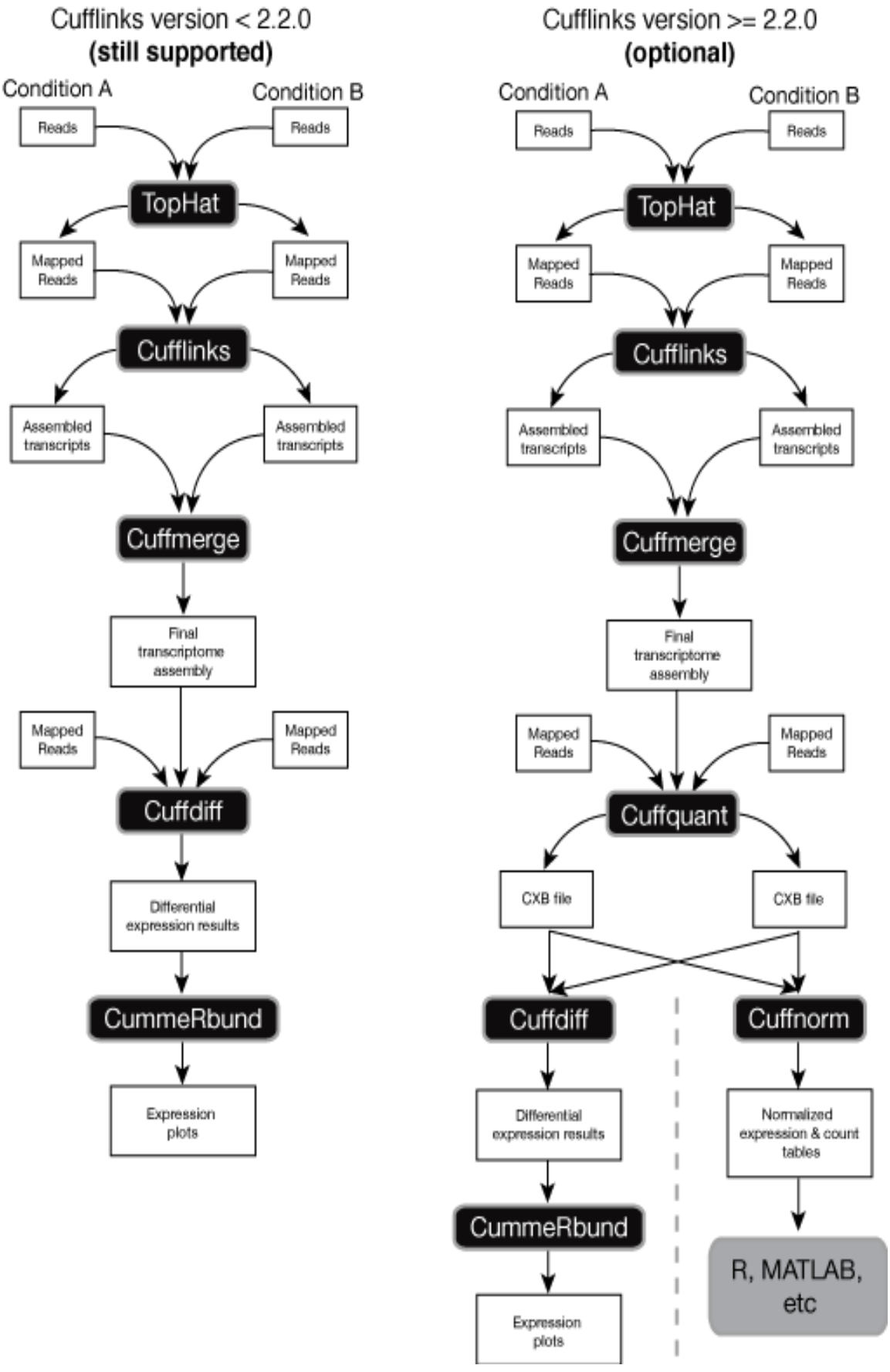
From the GSE53212 data entry 11 files were downloaded.4 biological replicates for Embryonic Stem Cell, 3 biological replicates for Induced Pluripotent StemCell, 2 biological replicates for Mouse Embryonic Fibroblast and 2 biological replicates for Nanog null Embryonic Stem cells.

The files obtained were in the **.sra** format.The **.sra** files were converted into **.fastq** or**.fasta** format by using the NCBI SRA TOOL KIT (windows x64)software which was downloaded from the NCBI DATABASE REPSITORY. Along with it Active Perl Configuration was downloaded to the same directory as NCBI TOOL KIT (sratoolkit.2.3.4-win64./bin) .Using command prompt remote connection was set with NCBI main database. The files were converted to **.fastq** format by the use of fastq-dump command.

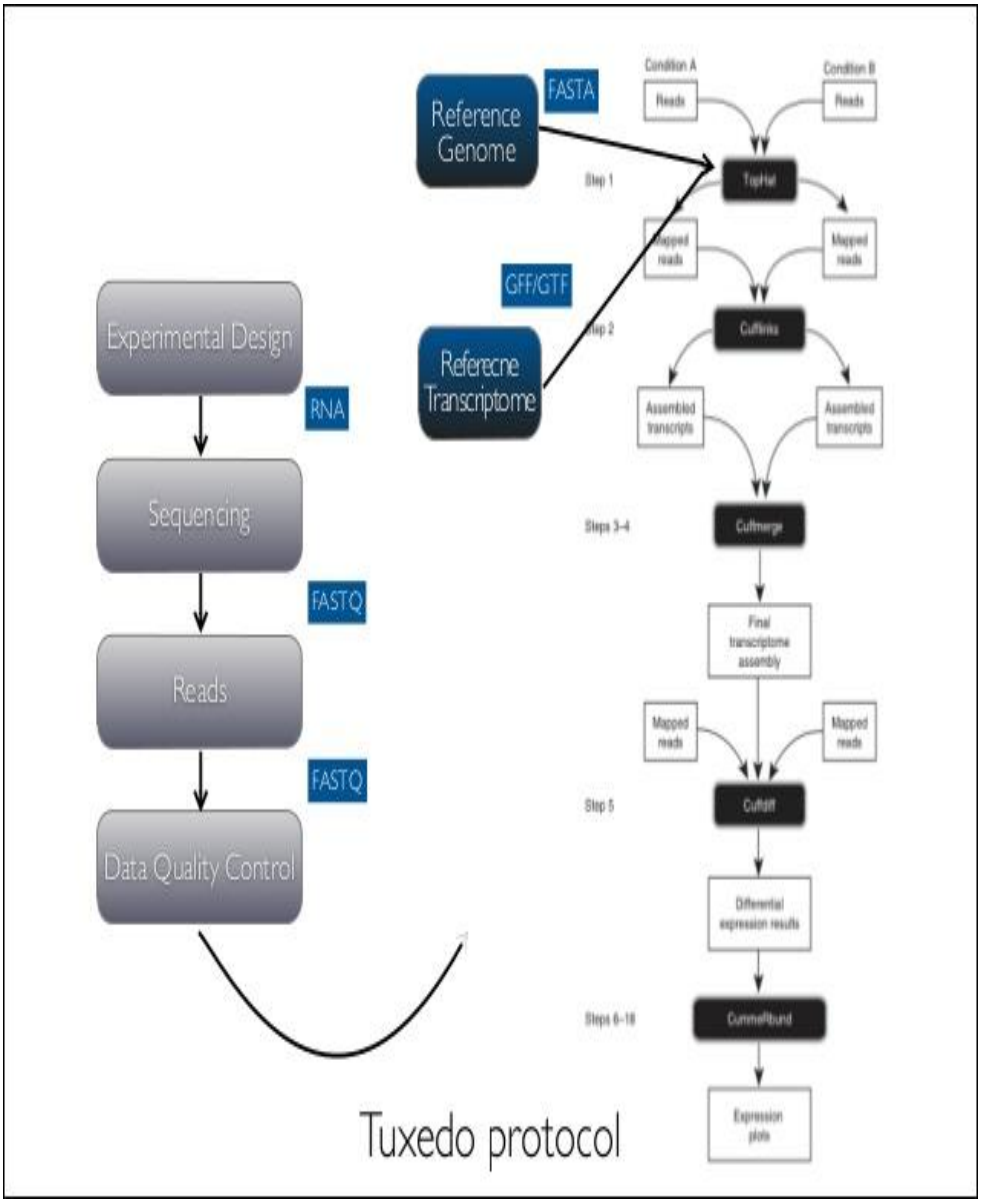
Syntax for the conversion of **.sra** file to **.fastq** file

C:\Users\USERS\Downloads\compressed\sratoolkit.2.3.4-win64\bin>fatsq-dump –A {filename} {file location} –O {destination folder}

The FASTQ files were then uploaded on to GALAXY via FTP client server Filezilla. GALAXY is an open source, web based platform for intensive biomedical and bioinformatics research. To identify differentially expressed genes and transcripts TUXEDO PROTOCOL was followed in GALAXY. The FASTQ files are imported onto the GALAXY Server by clicking on Get Data ->Upload . The raw sequence read file obtained from different high throughput sequencing technology can be converted to standard format by using FASTQ Groomer tool .Converting to standard format makes further downstream processing easier and less error prone. The quality of the raw sequence reads obtained from high throughput sequencing pipeline is calculated using GALAXY wrapper FASTQC. It provides a set of modular analyses about the quality of raw sequence reads. It accepts FASTQ ,SAM ,BAM file as input and provides a single HTML output file containing Basic statistics, Per base quality ,Per sequence quality scores, Per base sequence content, Per base GC content, Per sequence GC content, Per base N content, Sequence length distribution, Sequence length distribution, overrepresented sequences , Kimer content. FASTQ summary statistics tool is also used to obtain quality score statistics for each cycle. Box plot tool is used to determine whether the low median quality score needs to be removed. The reads with a phred quality score of greater than 20 are considered reliable for further analysis. The FASTQ files are trimmed using FASTQ Trimmer based on either Phred quality score of each base or based on the median quality score[20]. Phred quality score is calculated based on the noise produced by each base .Processed trimmed FASTQ files are used as input for the TUXEDO PROTOCOL.



(Trapnell et al.,2012)



TUXEDO PROTOCOL:

This protocol is a Bioinformatics protocol which is used to analyse the RNA-Seq data obtained from Next generation Sequencing techniques. The workflow is dependent on the use of various tools and softwares to optimize the reads obtained. The reads are mapped on to the reference genome if available else reference transcriptome is obtained by applying various Alignment algorithm.

TOPHAT(2):

For the Illumina reads which are provided by the input provided as input ,Tophat2 tool is used.Tophat2 is a splice junction mapper for RNA reads .It works in a two step process first ,it maps the reads to genome by ultra high throughput read aligner Bowtie(2) and in the second step it analyses common overlapping regions to determine splice junction between exons. The input format for Tophat(2) is Sanger FASTQ Format, Files are prepared by using FastQ Groomer .It provides output files pertaining to accepted_hits and junctions. Accepted _hits is a BAM file containing the list of read alignments and junctions is a BED track of junctions reported by TOPHAT.[18]

CUFFLINKS:

The accepted _hits BAM file generated by TOPHAT(2) is used as input to generate transcriptome assembly for each condition and to provide FPKM (Fragments per kilobase of exon) value estimation of each gene and transcript. Cufflinks assembles the transcripts from RNA-Seq which have been aligned to the genome. It is often not possible to predict all the splicing junction of a gene hence Cufflinks predicts a parsimonious transcriptome of assembled data. Cufflinks also works in a two step process first predicts the assembled transcripts and then it quantifies the expression of each transfag in the sample,it uses statistical sampling to exclude background or artifactual transfags (Trapnell et al.,2010). Cufflinks produces three output files transcripts.gtf, isoform.fpkm_tacking , genes.fpkm_tracking.[19]

CUFFMERGE:

This tool is a meta assembler, it treats the assembled transcriptome in the same way cufflinks treats mapped reads, merging them together parsimoniously. It follows reference annotation based transcriptome assembly(RABT) merges reference transcript with sample transfags to produce a file which is used to detect differential expression.[19]

CUFFDIFF:

Cuffdiff calculates the expression in two or more sample conditions and tests their statistical significance of each observed change.[19] The statistical model used to evaluate changes in the reads produced by transcripts is directly proportional to its abundance. Cuff diff provides eleven output files. The transcript and gene differential files can be imported and viewed in excel. Cuff diff is used to find significant changes in transcript expression, splicing ,primary transcript expression and promoter use. Cuffdiff takes Cufflink GTF file as input and two SAM files containing alignment of two or more replicates. The output files produced : Transcript FPKM expression tracking, Gene FPKM expression tracking; Primary transcript FPKM tracking; Coding sequence FPKM tracking, Transcript differential FPKM, Gene differential FPKM, Primary transcript differential FPKM, Coding sequence differential FPKM, Differential splicing tests, Differential promoter tests, Differential CDS tests.

CummeRbund:

CummeRbund is a tool designed to simplify the analysis, exploration and visualisation of RNA-Seq data derived from the output of differential expression analysis obtained by using Cuffdiff .[23]This tool is of prime importance in visualising multiple relation between various conditions and multiple replicates. Especially helpful in visualising appropriate

relationship between genes, transcripts, coding sequence regions, transcription start sites and junctional boundaries.

R: is a language used for statistical computing and graphics.It is a free software,which can be compiled and run on UNIX,WINDOWS and MacOS. R can be downloaded via CRAN MIRROR.[23]R provides a wide variety of statistical techniques including linear and non-linear modelling, real time analysis, time series and clustering along with wide variety of graphical techniques. R can be used to obtain publication ready graphs. R is preferred for visualising research data as it offers effective handling and storage of data, calculations are performed on arrays, presence of large number of coherent tools and functions for data analysis.R is a fully planned coherent system. The syntax for obtaining plots from R.

R can be downloaded from the source("http://bioconductor.org/bioclite.R")

```
biocLite()
biocLite("edgeR")
library(cummeRbund)
cuff<-readcufflinks()
disp<-dispersionplot(genes(cuff))
gsc<-fpkmscvplot(genes(cuff))
```

Database for Annotation, Visualization and Integrated Discovery (DAVID):

DAVID is a free source platform which provides tools for Functional Annotation, Gene functional Classification, Gene ID conversion, Gene Name Batch Viewer[22],[21].DAVID provides tools for identification of enriched biological themes, particularly Gene Ontology (GO) terms, Discovery of enriched functional-related gene groups, Clustering of redundant

annotation terms, Visualization of genes on BioCarta & KEGG pathway maps, Display of related many-genes-to-many-terms on 2-D view, Search for other functionally related genes not in the list , identification of interacting proteins, Exploration of gene names in batch ,Link gene-disease associations ,Highlighting protein functional domains and motifs .The list of differentially expressed transcripts was uploaded onto DAVID and the related Biological Process, Molecular function was obtained. It also provided the cellular location of the protein along with the related KEGG Pathways. DAVID can also be used to generate 2D heat maps for better visualisation of differentially expressed genes,it also provides rapid mean to reduce large lists of genes into functionally related cluster of genes.

CYTOSCAPE:

Cytoscape is an open source platform for visualizing molecular interaction networks and biological pathways and organising these networks with annotations, gene expression profiles and other state data.[24],[25]. Cytoscape is used for biological research, to analyse and visualize complex gene network. Cytoscape is used for integration, analysis and visualisation of genetic networks and cluster. Added features can be added by downloading Apps (Plugins) Apps are available for visualising molecular and genetic network , identifying and clustering molecular profiling , formation of new layouts, scripting, and connection with Public databases like UniProt , GeneMania. To identify the central gene in a molecular cluster and network, gene symbol were imported and were compared gene with 28 public databases. The genes are matched and gene networks and clustered are obtained. To obtain centre genes in cluster we used Uniprot, Reactome Netwrok and Reactome PI. Reactome Network App is used to detect the genes which are responsible and involved in cancer cell proliferation, migration, metastasis and invasion pathways. Reactome database used to determine the genes involved in molecular functions, summary of functional pathway ,to obtain network structure of the activated pathway.

CHAPTER FOUR: RESULTS AND DISCUSSION

RESULT AND DISCUSSION:

The 5 groups (Embryonic Stem Cell, Mouse Embryonic Fibroblast, Induced Pluripotent Stem Cell,Nanog null Induced Pluripotent Stem Cell,Nanog null Embryonic Stem Cell) each having 2-4 biological replicate. There are 16 transcripts common for the interaction between Nanog null Induced Pluripotent Stem Cell +iPSC and Embryonic Stem Cell + Induced Pluripotent Stem Cell , 17 transcripts common for the interaction between Nanog null Induced Pluripotent Stem Cell +iPSC Vs Mouse Embryonic Fibroblast+ Embryonic Stem Cell, 2 trancripts common for the interaction between Nanog null Embryonic Stem Cell+ Nanog null Induced Pluripotent Stem Cell Vs Embryonic Stem Cell+ Induced Pluripotent Stem Cell, 16 transcripts common for the interaction between Mouse Embryonic Fibroblast+ Embryonic Stem Cell Vs Nanog null Embryonic Stem Cell+ Nanog null Induced Pluripotent Stem Cell, 22 transcripts common for the interaction between Nanog null Embryonic Stem Cell+ Embryonic Stem Cell Vs Embryonic Stem Cell+ Induced Pluripotent Stem Cell, 22 transcripts common for the interaction between Nanog null Embryonic Stem Cell+ Embryonic Stem Cell Vs Embryonic Stem Cell Vs Mouse Embryonic Fibroblast+ Embryonic Stem Cell.

Comparative Interaction between groups	Number of common transcripts
Nanog(-/-)iPSC+iPSC Vs ESC+Ipsc	16
Nanog(-/-)iPSC+iPSC Vs MEF+ESC	17
Nanog(-/-)ESC+Nanog(-/-)iPSC Vs ESC+iPSC	02
MEF+ESC Vs Nanog(-/-)ESC+Nanog (-/-) iPSC	16
Nanog(-/-)ESC+ESC Vs ESC+iPSC	22
Nanog(-/-)ESC+ESC Vs MEF+ESC	22

The Data obtained from Cuff Diff was analysed and graphically visualised using R.The functional annotation of genes was done with the help of DAVID to retrieve information regarding Biological Processes , Molecular function and regarding cellular components.All the information along with FPKM value,Gene Symbol and locus was appended into data sheets.

Genes differential Expression for Nanog null iPSC Vs iPSC

Gene Sym	gene	sample_1	sample_2	value_1	value_2
Dst	NM_001276764,NM_010081,NM_1338	Nanog null ipsc	ipsc	7041.7	0
Dars	NM_145507,NM_177445	Nanog null ipsc	ipsc	30473	0
Ptprv	NM_007955	Nanog null ipsc	ipsc	10318.3	0
Jmjd1c	NM_001242396,NM_207221	Nanog null ipsc	ipsc	7391.74	0
Top2a	NM_011623	Nanog null ipsc	ipsc	12115.4	0
Nemf	NM_025441	Nanog null ipsc	ipsc	5.84E+08	1.25E+09
Gm2042	NM_001270792	Nanog null ipsc	ipsc	736470	511706
Extl3	NM_018788	Nanog null ipsc	ipsc	10502.5	0
Rpl8	NM_012053	Nanog null ipsc	ipsc	100051	0
Ep300	NM_177821	Nanog null ipsc	ipsc	7073.42	0
Ndrg1	NM_008681	Nanog null ipsc	ipsc	22843.2	0
Itga5	NM_010577	Nanog null ipsc	ipsc	14627.7	0
Eif4a2	NM_001123037,NM_001123038,NM_01	Nanog null ipsc	ipsc	20537.8	0
Gabbr1	NM_019439	Nanog null ipsc	ipsc	14212.1	0
Cyp1b1	NM_009994	Nanog null ipsc	ipsc	12337	0
Prdx1	NM_011034	Nanog null ipsc	ipsc	51975.3	10318
Caap1	NM_026368	Nanog null ipsc	ipsc	32502.3	0
Rbbp7	NM_009031	Nanog null ipsc	ipsc	29866.6	0

Gene Symbol	gene	sample_1	sample_2	value_1	value_2
Sgol2	NM_001177867,NM_199007	Nanog nul	iPSC	0	2938.64
Gls	NM_001081081,NM_001113383	Nanog nul	iPSC	0	2529.69
Rab1	NM_008996	Nanog nul	iPSC	0	4975.92
Prpf8	NM_138659	Nanog nul	iPSC	0	1650.83
Zfp207	NM_001130169,NM_001130170,NM_0011	Nanog nul	iPSC	0	5957.47
Nemf	NM_025441	Nanog nul	iPSC	5.84E+08	1.25E+09
Gm2042	NM_001270792	Nanog nul	iPSC	736470	511706
Apex1	NM_009687	Nanog nul	iPSC	0	22808.1
Rbfox2	NM_001110827,NM_001110828,NM_0011	Nanog nul	iPSC	0	1978.85
Rn45s	NR_046233	Nanog nul	iPSC	0	2139.18
Snx2	NM_026386	Nanog nul	iPSC	0	6806
Hdgf	NM_008231	Nanog nul	iPSC	0	6009.74
Prdx1	NM_011034	Nanog nul	iPSC	51975.3	10318
Elavl2	NM_001177883,NM_010486,NM_207685,I	Nanog nul	iPSC	0	3467.1
Gatad1	NM_026033	Nanog nul	iPSC	0	5215.46
Mcm7	NM_008568	Nanog nul	iPSC	0	5584.5
Pdap1	NM_001033313	Nanog nul	iPSC	0	6139.78
Plekha7	NM_172743	Nanog nul	iPSC	0	2942.56
Psma4	NM_011966	Nanog nul	iPSC	0	14289.2
Cmtm6	NM_026036	Nanog nul	iPSC	0	3856.75
Pnma5	NM_001100461	Nanog nul	iPSC	0	2.00E+07

Genes differential Expression for Nanog null iPSC Vs Nanog null ESC

Gene Symbol	gene	locus	sample_1	sample_2	value_1	value_2
Nek2	NM_010892	chr1:193645351-193656928	nanog nul	esc	21294.6	0
Degs1	NM_007853	chr1:184205900-184212890	nanog nul	esc	35857.4	0
Grn	NM_008175	chr11:102291635-102298123	nanog nul	esc	30178.5	0
Uqcr10	NM_197979	chr11:4601970-4604347	nanog nul	esc	349137	0
B4galt6	NM_019737	chr18:20843099-20904905	nanog nul	esc	11269.4	0
Got1	NM_010324	chr19:43574242-43599095	nanog nul	esc	34818.7	0
ubiquitin protein ligase E3 component n-recognin	NM_001081548,NM_177783	chr2:69735302-69862070	nanog nul	esc	7997.95	0
solute carrier family 20, member 1 (Slc20a1)	NM_001159593,NM_015747	chr2:129024508-129037348	nanog nul	esc	20552.4	0
Ncoa3	NM_008679	chr2:165818136-165898742	nanog nul	esc	8598.14	0
WW domain containing E3 ubiquitin protein lig	NM_001276292,NM_177327	chr4:19535445-19636151	nanog nul	esc	10835.6	0
Polr2b	NM_153798	chr5:77739508-77837070	nanog nul	esc	17605.6	0
Exoc4	NM_009148	chr6:33199149-33922930	nanog nul	esc	18126.4	0
Vamp8	NM_016794	chr6:72335214-72340661	nanog nul	esc	119272	0
Exoc6b	NM_177077	chr6:84568479-85019507	nanog nul	esc	12070.3	0
Capns1	NM_009795	chr7:30971960-30980067	nanog nul	esc	53509.7	0
Tbc1d10b	NM_144522	chr7:134340973-134351982	nanog nul	esc	18651.1	0
Map1s	NM_173013	chr8:73429872-73441428	nanog nul	esc	20750.2	0
Lars2	NM_153168	chr9:123276057-123371782	nanog nul	esc	17269.9	0
Cgnl1	NM_026599	chr9:71474315-71619409	nanog nul	esc	9627.21	0
Rps4x	NM_009094	chrX:99380281-99383710	nanog nul	esc	101789	0

Gene Symbol	gene	locus	sample_1	sample_2	value_1	value_2
Csrp2	NM_007792	chr10:110	nanog nul	esc	0	49839.5
Atp5b	NM_016774	chr10:127	nanog nul	esc	0	19325.8
Ubb	NM_011664	chr11:623	nanog nul	esc	0	34558.2
eukaryotic translation initiation fac	NM_001166589,NM_0011	chr11:697	nanog nul	esc	0	32751
Pfn1	NM_011072	chr11:704	nanog nul	esc	0	51396.9
eukaryotic translation initiation fac	NM_001005331,NM_145	chr16:206	nanog nul	esc	0	6073.05
Msh6	NM_010830	chr17:883	nanog nul	esc	0	7810.52
Scd2	NM_009128	chr19:443	nanog nul	esc	0	6022.34
COP9 (constitutive photomorphoge	NM_001285507,NM_0012	chr2:1256	nanog nul	esc	0	10897.3
Cpne3	NM_027769	chr4:1944	nanog nul	esc	0	5804.22
polypyrimidine tract binding protei	NM_144904,NM_178164	chr4:5948	nanog nul	esc	0	4730.96
Uso1	NM_019490	chr5:9256	nanog nul	esc	0	8594.58
Mlec	NM_175403	chr5:1155	nanog nul	esc	0	5492.08
general transcription factor II I (Gtf2	NM_001080746,NM_0010	chr5:1347	nanog nul	esc	0	7656.79
Gabarapl1	NM_020590	chr6:1294	nanog nul	esc	0	20681.5
Pole4	NM_025882	chr6:8259	nanog nul	esc	0	21822.4
Slc2a3	NM_011401	chr6:1226	nanog nul	esc	0	8449.89
protein kinase, cAMP dependent, c	NM_001277898,NM_0088	chr8:8649	nanog nul	esc	0	15735.1
Asna1	NM_019652	chr8:8754	nanog nul	esc	0	31533.4
Bgn	NM_007542	chrX:7072	nanog nul	esc	0	14692

Gene differential Expression for Nanog null ESC Vs ESC

Gene Symbol	gene	locus	sample_1	sample_2	value_1	value_2
Csrp2	NM_007792	chr10:110	nanog nul	esc	0	49839.5
Atp5b	NM_016774	chr10:127	nanog nul	esc	0	19325.8
Ubb	NM_011664	chr11:623	nanog nul	esc	0	34558.2
eukaryotic translation initiation fac	NM_001166589,NM_0011	chr11:697	nanog nul	esc	0	32751
Pfn1	NM_011072	chr11:704	nanog nul	esc	0	51396.9
eukaryotic translation initiation fac	NM_001005331,NM_145	chr16:206	nanog nul	esc	0	6073.05
Msh6	NM_010830	chr17:883	nanog nul	esc	0	7810.52
(D17Wsu92e)	NM_001033279,NM_0010	chr17:278	nanog nul	esc	0	8799.67
bromodomain containing 2 (Brd2)	NM_001204973,NM_0102	chr17:342	nanog nul	esc	0	7710.5
Scd2	NM_009128	chr19:443	nanog nul	esc	0	6022.34
COP9 (constitutive photomorphoge	NM_001285507,NM_0012	chr2:1256	nanog nul	esc	0	10897.3
Cpne3	NM_027769	chr4:1944	nanog nul	esc	0	5804.22
polypyrimidine tract binding protei	NM_144904,NM_178164	chr4:5948	nanog nul	esc	0	4730.96
Uso1	NM_019490	chr5:9256	nanog nul	esc	0	8594.58
Mlec	NM_175403	chr5:1155	nanog nul	esc	0	5492.08
general transcription factor II I (Gtf2	NM_001080746,NM_0010	chr5:1347	nanog nul	esc	0	7656.79
Gabarapl1	NM_020590	chr6:1294	nanog nul	esc	0	20681.5
Pole4	NM_025882	chr6:8259	nanog nul	esc	0	21822.4
Slc2a3	NM_011401	chr6:1226	nanog nul	esc	0	8449.89
protein kinase, cAMP dependent, c	NM_001277898,NM_0088	chr8:8649	nanog nul	esc	0	15735.1
Asna1	NM_019652	chr8:8754	nanog nul	esc	0	31533.4
Bgn	NM_007542	chrX:7072	nanog nul	esc	0	14692

Gene Symbol	gene	locus	sample_1	sample_2	value_1	value_2
Nek2	NM_010892	chr1:193645351-193656928	nanog nul	esc	21294.6	0
Degs1	NM_007853	chr1:184205900-184212890	nanog nul	esc	35857.4	0
Fnip1	NM_173753	chr11:54251680-54331743	nanog nul	esc	10383.8	0
Grn	NM_008175	chr11:102291635-102298123	nanog nul	esc	30178.5	0
Uqcr10	NM_197979	chr11:4601970-4604347	nanog nul	esc	349137	0
Ubald1	NM_145359	chr16:4874778-4879851	nanog nul	esc	55604.3	0
Rn45s	NR_046233	chr17:39979941-39985774	nanog nul	esc	11218.8	0
B4galt6	NM_019737	chr18:20843099-20904905	nanog nul	esc	11269.4	0
Got1	NM_010324	chr19:43574242-43599095	nanog nul	esc	34818.7	0
ubiquitin protein ligase E3 component n-recognin	NM_001081548,NM_177783	chr2:69735302-69862070	nanog nul	esc	7997.95	0
solute carrier family 20, member 1 (Slc20a1), tr	NM_001159593,NM_015747	chr2:129024508-129037348	nanog nul	esc	20552.4	0
Ncoa3	NM_008679	chr2:165818136-165898742	nanog nul	esc	8598.14	0
Lurap1l	NM_026821	chr4:80556589-80600205	nanog nul	esc	43493	0
WW domain containing E3 ubiquitin protein lig	NM_001276292,NM_177327	chr4:19535445-19636151	nanog nul	esc	10835.6	0
Polr2b	NM_153798	chr5:77739508-77837070	nanog nul	esc	17605.6	0
Exoc4	NM_009148	chr6:33199149-33922930	nanog nul	esc	18126.4	0
Vamp8	NM_016794	chr6:72335214-72340661	nanog nul	esc	119272	0
Exoc6b	NM_177077	chr6:84568479-85019507	nanog nul	esc	12070.3	0
Capns1	NM_009795	chr7:30971960-30980067	nanog nul	esc	53509.7	0
Tbc1d10b	NM_144522	chr7:134340973-134351982	nanog nul	esc	18651.1	0
Map1s	NM_173013	chr8:73429872-73441428	nanog nul	esc	20750.2	0
Tmem231	NM_001033321	chr8:114435917-114457691	nanog nul	esc	23809.4	0

Genes differential Expression for ESC Vs iPSC

gene syml	gene	locus	sample_1	sample_2	value_1	value_2
Csrp2	NM_0077	chr10:110	ESC	iPsc	44301.8	0
Atp5b	NM_0167	chr10:127	ESC	iPsc	17178.4	0
Ubb	NM_0116	chr11:623	ESC	iPsc	110586	0
Eif5a	NM_0011	chr11:697	ESC	iPsc	29112.8	0
Pfn1	NM_0110	chr11:704	ESC	iPsc	45686.2	0
Eif4g1	NM_0010	chr16:206	ESC	iPsc	19459.7	0
Msh6	NM_0108	chr17:883	ESC	iPsc	6942.69	0
D17Wsu92	NM_0010	chr17:278	ESC	iPsc	28158.9	0
Brd2	NM_0012	chr17:342	ESC	iPsc	6854.02	0
Scd2	NM_0091	chr19:443	ESC	iPsc	5353.19	0
Sgn2	NM_0012	chr2:1256	ESC	iPsc	34875.2	0
Cpne3	NM_0277	chr4:1944	ESC	iPsc	5159.31	0
Ptbp3	NM_1449	chr4:5948	ESC	iPsc	4205.3	0
Uso1	NM_0194	chr5:9256	ESC	iPsc	7639.63	0
Mlec	NM_1754	chr5:1155	ESC	iPsc	4881.85	0
Gtf2i	NM_0010	chr5:1347	ESC	iPsc	6807.09	0
Gabarapl1	NM_0205	chr6:1294	ESC	iPsc	18383.5	0
Pole4	NM_0258	chr6:8259	ESC	iPsc	19397.7	0
Slc2a3	NM_0114	chr6:1226	ESC	iPsc	27039.7	0
Prkaca	NM_0012	chr8:8649	ESC	iPsc	13988	0
Asna1	NM_0196	chr8:8754	ESC	iPsc	28029.7	0
Bgn	NM_0075	chrX:7072	ESC	iPsc	13059.5	0

gene syml	gene	locus	sample_1	sample_2	status	value_1	value_2
Nemf	NM_025441	chr12:702(	ESC	iPsc	OK	0	9.19E+08
Pnma5	NM_001100461	chrX:7027	ESC	iPsc	OK	0	1.47E+07
Gm2042	NM_001270792	chr12:891	ESC	iPsc	NOTEST	0	376647
Tdgf1	NM_011562	chr9:1108	ESC	iPsc	NOTEST	0	42883.5
Stip1	NM_016737	chr19:709	ESC	iPsc	NOTEST	0	37365
Gm13157	NM_001127189	chr4:1471	ESC	iPsc	NOTEST	0	28946.1
Bag6	NM_001252468,NM_001252469,NM_057171	chr17:352	ESC	iPsc	NOTEST	0	21597.1
Apex1	NM_009687	chr14:515	ESC	iPsc	NOTEST	0	16788.2
Hspa4	NM_008300	chr11:530	ESC	iPsc	NOTEST	0	16312.9
Ppfia4	NM_001144855	chr1:1361	ESC	iPsc	NOTEST	0	12303.3
Psma4	NM_011966	chr9:5479	ESC	iPsc	NOTEST	0	10517.7
Prdx1	NM_011034	chr4:1163	ESC	iPsc	NOTEST	0	7594.66
Snx2	NM_026386	chr18:533	ESC	iPsc	NOTEST	0	5009.63
Pdap1	NM_001033313	chr5:1458	ESC	iPsc	NOTEST	0	4519.25
Hdgf	NM_008231	chr3:8771	ESC	iPsc	NOTEST	0	4423.54
Zfp207	NM_001130169,NM_001130170,NM_00113017	chr11:801	ESC	iPsc	NOTEST	0	4385.18
Mcm7	NM_008568	chr5:1386	ESC	iPsc	NOTEST	0	4110.53
Gatad1	NM_026033	chr5:3639	ESC	iPsc	NOTEST	0	3838.9
Rab1	NM_008996	chr11:201	ESC	iPsc	NOTEST	0	3662.58
Cmtm6	NM_026036	chr9:1146	ESC	iPsc	NOTEST	0	2838.8
Elavl2	NM_001177883,NM_010486,NM_207685,NM_	chr4:9091	ESC	iPsc	NOTEST	0	2552.2
Plekha7	NM_172743	chr7:1232	ESC	iPsc	NOTEST	0	2165.9

Genes differential Expression for iPSC Vs MEF

gene symbol	gene	locus	sample_1	sample_2	value_1	value_2
Nemf	NM_025441	chr12:702	MEF	ipsc	0	9.19E+08
Pnma5	NM_001100461	chrX:7027	MEF	ipsc	0	1.47E+07
Gm2042	NM_001270792	chr12:891	MEF	ipsc	0	376647
Tdgf1	NM_011562	chr9:1108	MEF	ipsc	0	42883.5
Stip1	NM_016737	chr19:709	MEF	ipsc	0	37365
Gm13157	NM_001127189	chr4:1471	MEF	ipsc	0	28946.1
Bag6	NM_001252468,NM_001252469,NM_057171	chr17:352	MEF	ipsc	0	21598.3
Apex1	NM_009687	chr14:515	MEF	ipsc	0	16788.2
Hspa4	NM_008300	chr11:530	MEF	ipsc	0	16312.9
Ppfia4	NM_001144855	chr1:1361	MEF	ipsc	0	12303.3
Psma4	NM_011966	chr9:5479	MEF	ipsc	0	10517.7
Prdx1	NM_011034	chr4:1163	MEF	ipsc	0	7594.66
Snx2	NM_026386	chr18:533	MEF	ipsc	0	5009.63
Pdap1	NM_001033313	chr5:1458	MEF	ipsc	0	4519.25
Hdgf	NM_008231	chr3:8771	MEF	ipsc	0	4423.54
Zfp207	NM_001130169,NM_001130170,NM_00113017	chr11:801	MEF	ipsc	0	4386.69
Mcm7	NM_008568	chr5:1386	MEF	ipsc	0	4110.53
Gatad1	NM_026033	chr5:3639	MEF	ipsc	0	3838.9
Rab1	NM_008996	chr11:201	MEF	ipsc	0	3662.58
Cmtm6	NM_026036	chr9:1146	MEF	ipsc	0	2838.8
Elavl2	NM_001177883,NM_010486,NM_207685,NM_	chr4:9091	MEF	ipsc	0	2554.18
Plekha7	NM_172743	chr7:1232	MEF	ipsc	0	2165.9

gene symbol	gene	locus	sample_1	sample_2	value_1	value_2
smg7	NM_001005507,NM_001160256,NM_0(	chr1:1546	MEF	ipsc	7733.4	0
Ppp2ca	NM_019411	chr11:519	MEF	ipsc	24961	0
Tvp23b	NM_026210	chr11:626	MEF	ipsc	56031	0
canx	NM_001110499,NM_001110500,NM_0(	chr11:501(	MEF	ipsc	10424.8	0
sparc	NM_001290817,NM_009242	chr11:552(	MEF	ipsc	17516.1	0
Vat1	NM_012037	chr11:101	MEF	ipsc	16860.4	0
trim47	NM_001205081,NM_172570	chr11:115	MEF	ipsc	21558	0
Hnrnpa1	NM_001039129,NM_010447	chr15:103(	MEF	ipsc	36961.6	0
Zfp251	NM_001007568	chr15:766	MEF	ipsc	12859	0
3-hydroxybutyra	NM_001122683,NM_175177	chr16:314	MEF	ipsc	14660.3	0
Abcf1	NM_013854	chr17:360	MEF	ipsc	14421.2	0
Csf1r	NM_001037859	chr18:612(	MEF	ipsc	11528.4	0
Thbs1	NM_011580	chr2:1179	MEF	ipsc	7519.33	0
Notch1	NM_008714	chr2:2631	MEF	ipsc	4515.11	0
Rcn1	NM_009037	chr2:1052	MEF	ipsc	14868.9	0
Ptx3	NM_008987	chr3:6585	MEF	ipsc	25081.2	0
clathrin, light pc	NM_001080384,NM_001080385,NM_0(	chr4:4402	MEF	ipsc	51154.1	0
Spsb1	NM_029035	chr4:1492	MEF	ipsc	14916.8	0
Abcg2	NM_011920	chr6:5854(	MEF	ipsc	18682.3	0
Raf1	NM_029780	chr6:1155	MEF	ipsc	14853	0
WNK lysine defi	NM_001185020,NM_001185021,NM_0(	chr6:1198	MEF	ipsc	4397.32	0
Urb2	NM_001029876	chr8:1265	MEF	ipsc	7558.89	0

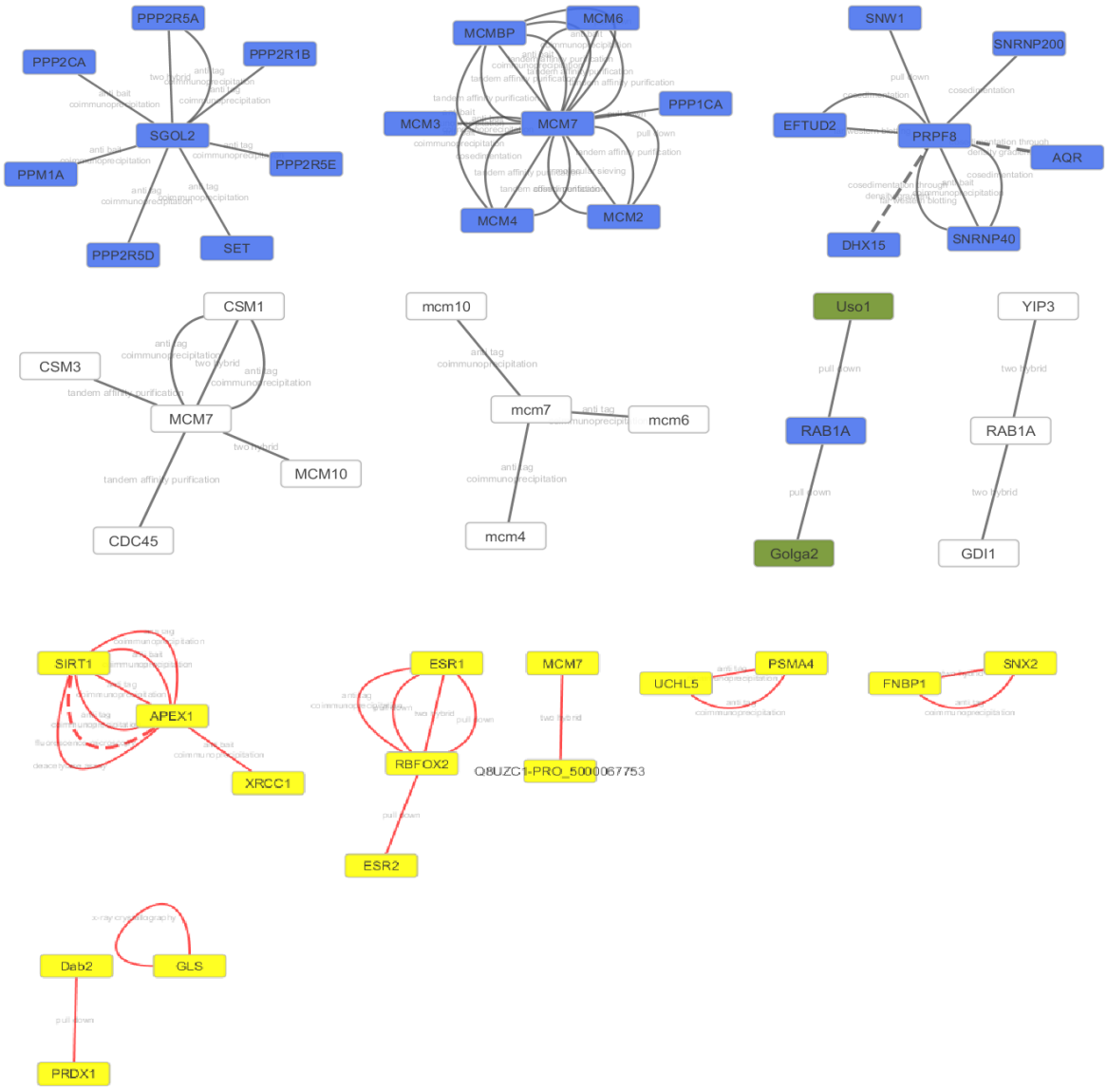
Gene differential Expression for ESC Vs MEF

gene name	gene	locus	sample_1	sample_2	value_1	value_2
Dst	NM_001276764,NM_010081,NM_13383	chr1:3396	MEF	ESC	7330.92	0
Dars	NM_145507,NM_177445	chr1:1302(	MEF	ESC	31725.7	0
Ptprv	NM_007955	chr1:1370(	MEF	ESC	10742.5	0
Jmjd1c	NM_001242396,NM_207221	chr10:665	MEF	ESC	7693.62	0
Top2a	NM_011623	chr11:988	MEF	ESC	12613.5	0
Extl3	NM_018788	chr14:656	MEF	ESC	10934.2	0
Rpl8	NM_012053	chr15:767	MEF	ESC	104164	0
Ep300	NM_177821	chr15:814	MEF	ESC	7364.19	0
Ndrg1	NM_008681	chr15:667(	MEF	ESC	23782.2	0
Itga5	NM_010577	chr15:103	MEF	ESC	15229	0
Eif4a2	NM_001123037,NM_001123038,NM_01	chr16:231(	MEF	ESC	21382	0
Gabbr1	NM_019439	chr17:371	MEF	ESC	14796.4	0
Cyp1b1	NM_009994	chr17:801(	MEF	ESC	12844.2	0
Prdx1	NM_011034	chr4:1163	MEF	ESC	54111.9	0
Caap1	NM_026368	chr4:9416(	MEF	ESC	33838.4	0
Rbbp7	NM_009031	chrX:1591	MEF	ESC	31094.3	0

gene name	gene	locus	sample_1	sample_2	value_1	value_2
Csrp2	NM_007792	chr10:110357231-110	MEF	ESC	0	46907.7
Atp5b	NM_016774	chr10:127520362-127	MEF	ESC	0	18188.9
Ubb	NM_011664	chr11:62365005-6236	MEF	ESC	0	110586
Eif5a	NM_001166589,NM_001166590,I	chr11:69730213-6973	MEF	ESC	0	30822.7
Pfn1	NM_011072	chr11:70465348-7046	MEF	ESC	0	48373.6
Eif4g1	NM_001005331,NM_145941	chr16:20672821-2069	MEF	ESC	0	19440.3
Msh6	NM_010830	chr17:88374389-8846	MEF	ESC	0	7351.08
D17Wsu92e	NM_001033279,NM_001044719,I	chr17:27888176-2795	MEF	ESC	0	28158.9
Brd2	NM_001204973,NM_010238	chr17:34248963-3425	MEF	ESC	0	7256.37
Scd2	NM_009128	chr19:44368165-4438	MEF	ESC	0	5668.08
Cops2	NM_001285507,NM_001285512,I	chr2:125656037-1256	MEF	ESC	0	34874.4
Cpne3	NM_027769	chr4:19446398-19497	MEF	ESC	0	5462.8
Ptbp3	NM_144904,NM_178164	chr4:59484739-59562	MEF	ESC	0	4452.67
Uso1	NM_019490	chr5:92566963-92631	MEF	ESC	0	8089.02
Mlec	NM_175403	chr5:115592989-1156	MEF	ESC	0	5169.02
Gtf2i	NM_001080746,NM_001080747,I	chr5:134713701-1347	MEF	ESC	0	7204.13
Gabarapl1	NM_020590	chr6:129482201-1294	MEF	ESC	0	19464.9
Pole4	NM_025882	chr6:82596705-82602	MEF	ESC	0	20538.8
Slc2a3	NM_011401	chr6:122677826-1226	MEF	ESC	0	27039.7
Prkaca	NM_001277898,NM_008854	chr8:86496876-86520	MEF	ESC	0	14806.8
Asna1	NM_019652	chr8:87541829-87549	MEF	ESC	0	29678.5
Bgn	NM_007542	chrX:70728973-70741	MEF	ESC	0	13827.7

Using Cytoscape we were able to identify genes in every comparative interaction which were played a pivotal role in any cellular and molecular pathways. To compare the change we calculated the relative expression of the enriched gene in the control cell lines as well as in the Nanog null cell lines.

For the comparative interaction between Nanog(-/-)iPSC+iPSC Vs ESC+iPSC, the enriched genes were Sgol2,Mcm7,Prpf8,Psma4,Apex1.

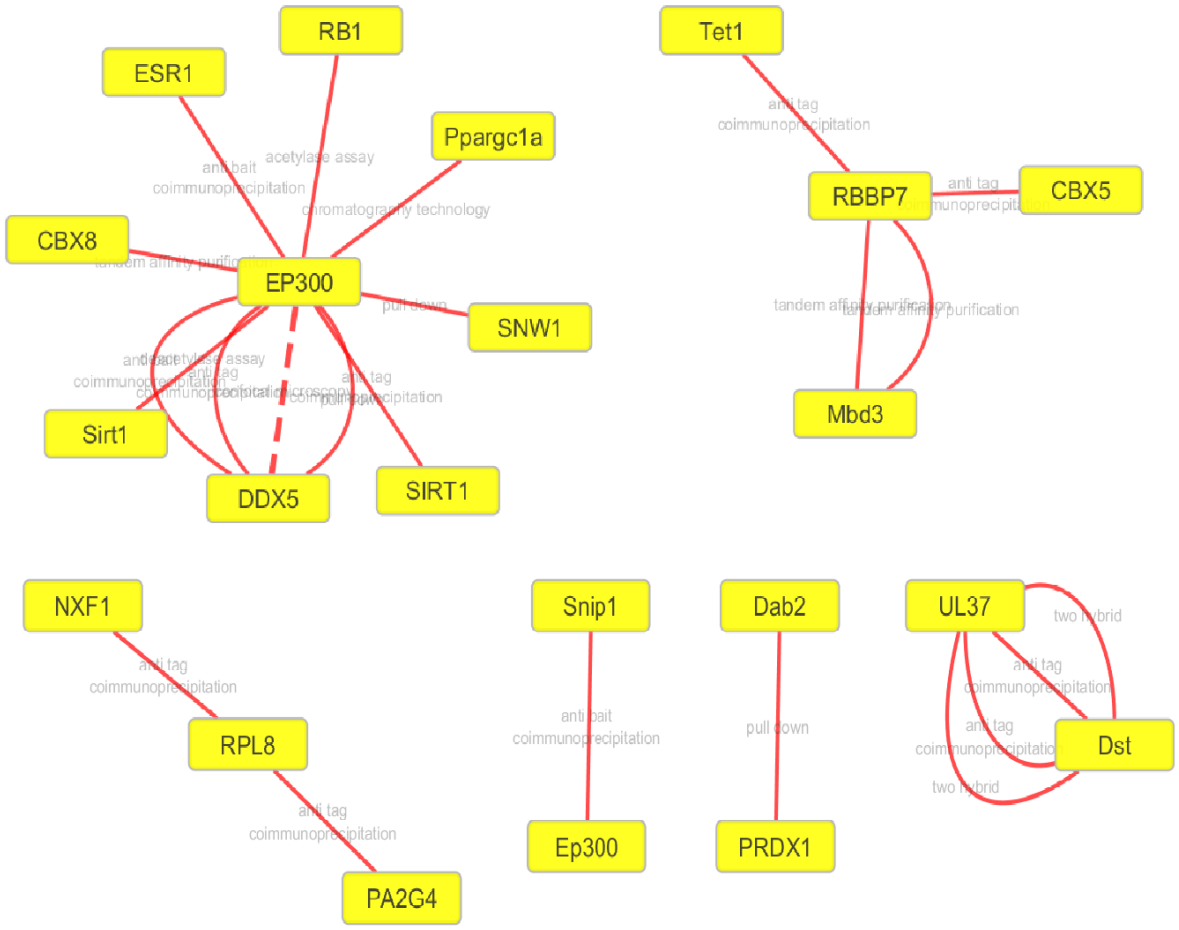


RELATIVE	FPKM value in 873	FPKM value in 918	Relative Expression
Sgol2	2162.14	2937.4	1.358561425
Mcm7	4110.53	5584.5	1.35858393
Prpf8	1215.11	1215.11	1
Psma4	10517.7	14289.2	1.358586003

Apex1	167882.2	22808.1	0.135857762

[RELATIVE EXPRESSION =FPKM VALUE IN 918/FPKM VALUE IN 873]

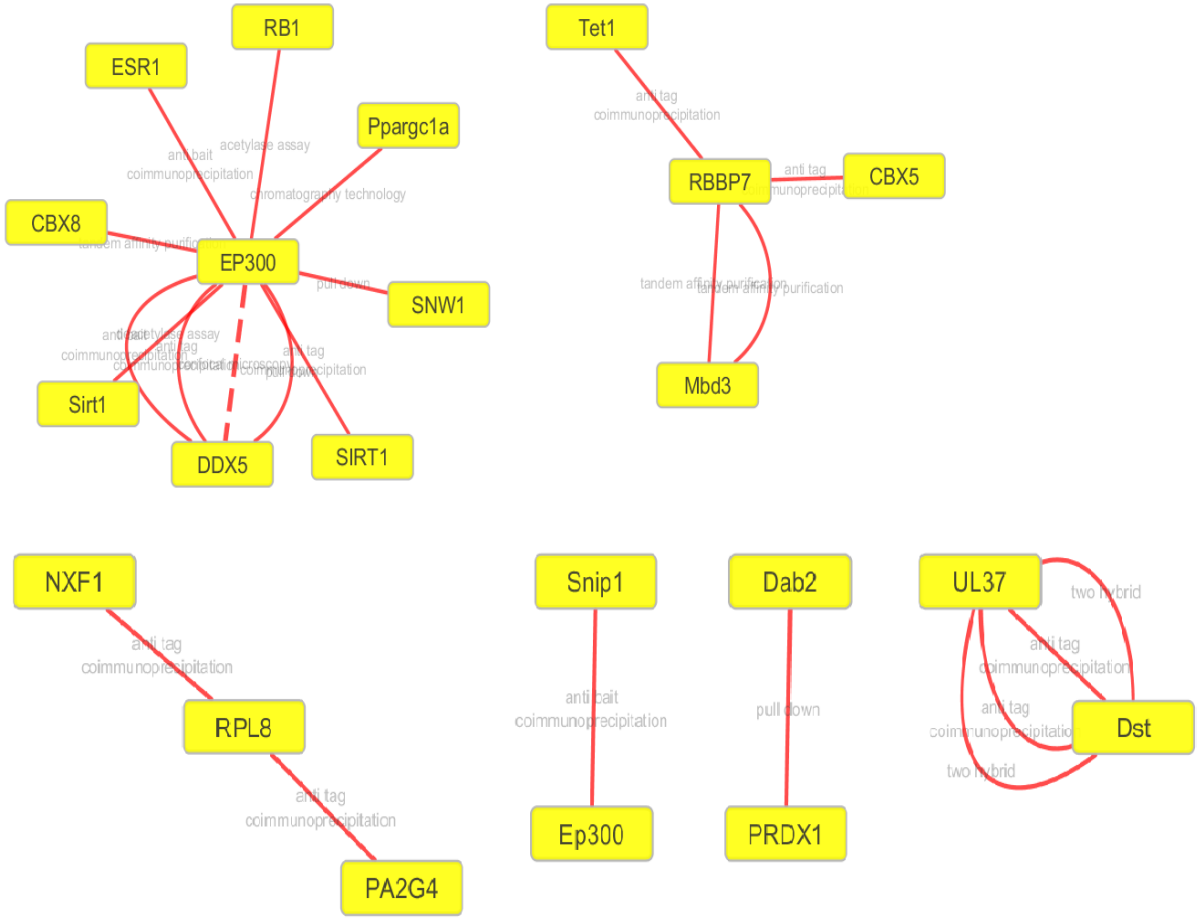
For the comparative interaction between Nanog(-/-)iPSC+iPSC Vs MEF+ESC the enriched genes were Ep300,Rbbp7,Rpl8,Dars.



Gene Symbol	FPKM value in 858	FPKM value in 918	Relative Expression
Ep300	23782.2	7073.42	0.297424965
Rbbp7	31094.3	29866.6	0.960516879
Rpl8	104164	100051	0.960514189
Dars	31725.7	8047.3	0.253652402

[RELATIVE EXPRESSION=FPKM VALUE IN 918/FPKM VALUE IN 858]

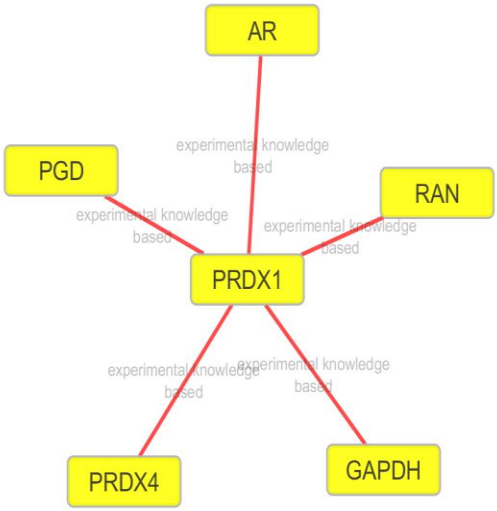
For the interaction between MEF+ESC Vs Nanog(-/-)ESC+Nanog (-/-) iPSC the enriched genes are Ep300,Rbbp7,Rpl8,Itag5,Eif4a2.



Gene Symbol	FPKM value in 858	FPKM value in 813	Relative Expression
Ep300	23782.2	7364.19	0.309651336
Rbbp7	31094.3	31094.3	1
Rpl8	104164	104164	1
Itag5	15229	15229	1
Eif4a2	21382	21382	1

[RELATIVE EXPRESSION=FPKM VALUE IN 918/FPKM VALUE IN 858]

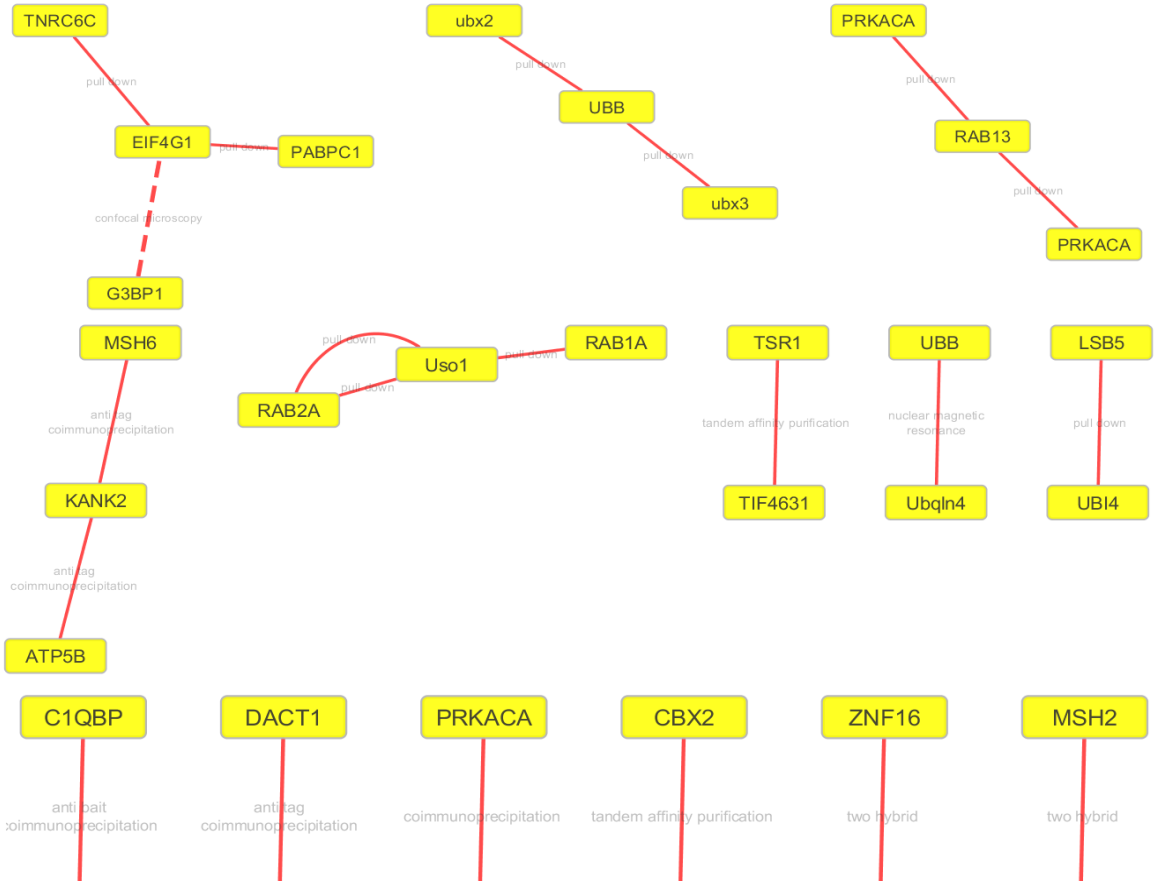
For the interaction between Nanog null ESC+Nanog null iPSC Vs ESC+iPSC, common genes were Rn45s,Prdx1.And Prdx1 is the only enriched gene.



Gene Symbol	FPKM value in 813	FPKM value in 873	Relative expression
Prdx1	54111.9	7594.66	7.124993087

[RELATIVE EXPRESSION=FPKM VALUE IN 873/FPKM VALUE IN 813]

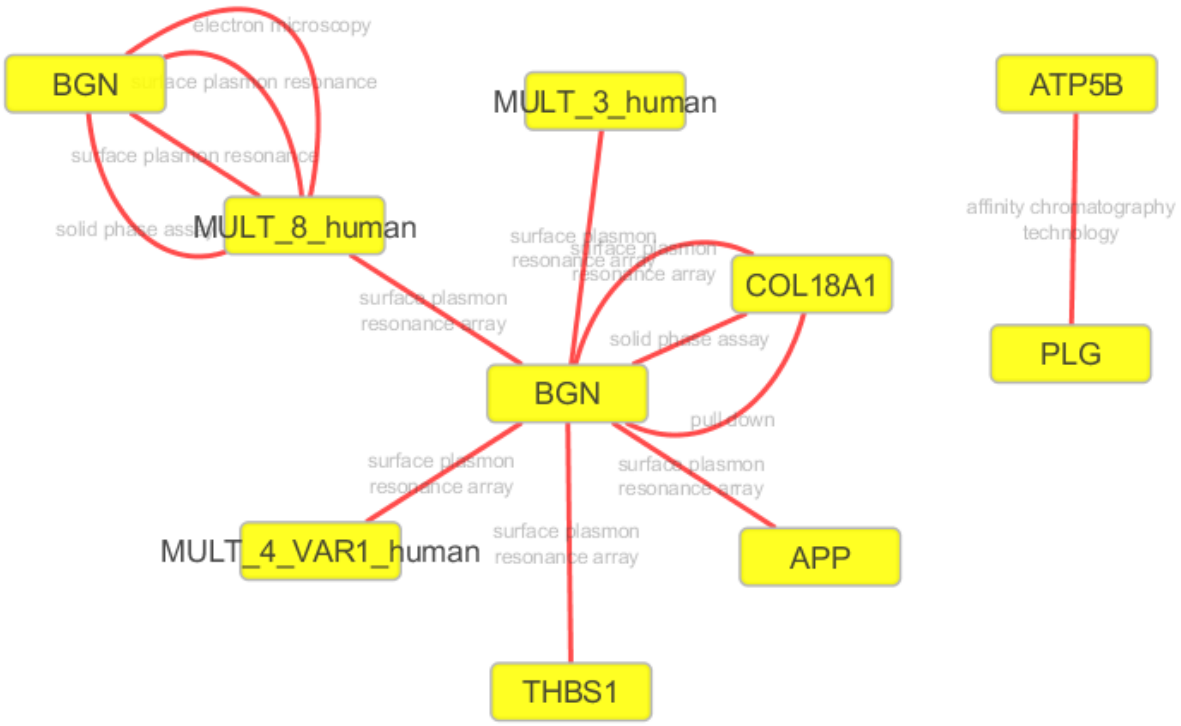
For the interaction between Nanog(-/-)ESC+ESC Vs ESC+iPSC the genes enriched were Eif4g1,Ubb,Rab13..



Gene Symbol	FPKM value in 858	FPKM value in 937	Relative Expression
Eif4g1	3299.13	3061.96	0.928111351
Prkaca	14342	15370.3	1.071698508
Cops2	34846.3	10875.3	0.312093393
Uso1	8089.02	8594.598	1.062501762

[RELATIVE EXPRESSION=FPKM VALUE IN 937/FPKM VALUE IN 858]

For the interaction between Nanog(-/-)ESC+ESC Vs MEF+ESC Eif4g1,Prkaca,Cops2,Uso1



Gene Symbol	FPKM value in 858	FPKM value in 937	Relative Expression
Eif4g1	3299.13	3061.96	0.928111351
Prkaca	14342	15370.3	1.071698508
Cops2	34846.3	10875.3	0.312093393
Uso1	8089.02	8594.598	1.062501762

[RELATIVE EXPRESSION=FPKM VALUE IN 937/FPKM VALUE IN 858]

Taking into consideration all the genes which were enriched Sgol2,Mcm7,Psma4,Prakca,,USo1 and Prdx1 are up-regulated in the Nanog null cell lines (Having Relative Expression >1.1).And Apex1,Ep300,Rbbp7,Rpl8,Dars,Eif4gi,Ubb,Cops2 are down regulated in Nanog null Cell lines(Having Relative Expression<1).Mildly up-regulated (1<=Relative Expression<1.1) Prkaca,Uso1,Eif4a1 The use of Tuxedo protocol followed by the functional annotation in DAVID and identifying cluster networks through the use of Cytoscape, provided a better understanding of the Nanog dependent pathway. The

genes which were found to be upregulated are Sgol2,Mcm7 and Psma4 ,all the genes are involved in the regulation of cell cycle, associated with chromatid cohesion and associated with pre-replication process. Sgol2 associated with G1/S transition,DNA replication.And the genes which have been down regulated act as antagonist to cell replication .Apex 1 encodes for enzyme DNA apurinic lyase , which prevents normal DNA replication.Ep300 is associated with the process of cellular growth proliferation and differentiation.Rbbp7 is Retinoblastoma binding protein ,it regulates cell proliferation and differentiation and is known to be associated withBRAC1 tumour suppressor gene.Rpl8 and Dars are genes which are associated with the process of protein synthesis. Ubb also known as Ubiquitin B is associated with the maintenance of chromatin structure, the regulation of gene expression, and the stress response. Cops2 is involved in early stages of neuronal differentiation ,involved in various cellular and developmental process.

CONCLUSION

The Data Set GSE53212 was analysed using Tuxedo Protocol, followed by functional annotation done by DAVID and further determining gene cluster using Cytoscape. Results showed that, in Nanog null cell lines, Sgol2, Mcm7, Psma4 were up-regulated by a factor of 1.36 whereas several pluripotency related genes were down regulated. The up-regulated genes were associated with the regulation of cell cycle and regulation of pre replication complex whereas down-regulated genes were found to be associated with protein synthesis pathways, cellular proliferation and differentiation pathways besides pluripotency. Further, upon Nanog knock-down, genes associated with the process of replication were activated while genes associated with the cellular differentiation were down-regulated by a relative expression in the range of (0.3-0.9). In conclusion, the results of the present study suggests that Nanog may not only be responsible for determining the pluripotency and reprogramming of cells but also for the process of cellular replication .

REFERENCES:

[1]Wang Zhong,Gerstein Mark and Snyder Michael. RNA-Seq a revolutionary tool for trnacriptomics, Nat Rev Genet. Jan 2009; 10(1): 57–63.

[2]Martin A.Jeffrey and Wang Zhong. Next-generation transcriptome assembly , Nature Reviews Genetics 12, (October 2011) ,671-682.

[3]Florea D. Lillian and Salzberg L. Steven. Genome-Guided Transcriptome Assemblyin the Age of Next-Generation Sequencing, Transcations on Computational Biology and Bioinofmratics,Vol.10,NO.5,septembet /October 2013.

[4] Chambers I., Colby D., Robertson M., Nichols J., Lee S., Tweedie S., Smith A. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell. 2003;113:643–655.

[5] Monya Baker. What does Nanog do?. Nature Reports Stem Cells, Nature Reports Stem Cells.

[6] Silva, J. et al. Nanog is the gateway to the pluripotent ground state. Cell 138, (2009), 722–737 .

[7]Carter C. Ava,Brandi N. Davis-Dusenbury,Kathryn Koszka,Ichida K. Justin and Eggan Kevin. Nanog-Independent Reprogramming to iPSCs with Canonical Factors.Stem Cell Reports, Volume 2, Issue 2, , 11 February 2014, p119–126.

[8] Moon JH1, Yun W, Kim J, Hyeon S, Kang PJ, Park G, Kim A, Oh S, Whang KY, Kim DW, Yoon BS and You S. Reprogramming of mouse fibroblasts into induced pluripotent stem cells with Nanog, Biochem Biophys Res Commun. 2013 Feb 15;431(3).

[9] Mitsui K1, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M and Yamanaka S. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell. 2003 May 30;113(5):631-42.

[10] Tongxiang Lin1, Connie Chao1, Shin'ichi Saito2, Sharlyn J. Mazur2, Maureen E. Murphy3, Ettore Appella2 and Yang Xu1. p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. Nature Cell Biology 7, (2004) ,165-171.

[11] Hamazaki takashi,Kehoe M.Sarah and Terada Naohiro. The Grb2/Mek Pathway Represses Nanog in Murine Embryonic Stem Cells. Mol Cell Biol,Oct 2006, 26(20): 7539–7549.

[12] Chiou SH, Yu CC and Huang CY.Positive correlations of Oct-4 and Nanog in oral cancer stem-like cells and high-grade oral squamous cell carcinoma. Clin Cancer Res,2008,4:4085–4095.

[13] Bussolati B, Bruno S and Grange C. Identification of a tumor-initiating stem cell population in human renal carcinomas FASEB J.,2008,22:3696–3675.

[14] Chambers I, Silva J and Colby D.Nanog safeguards pluripotency and mediates germline development, Nature. 2007,450:1230–1234.

[15] Hart AH, Hartley L and Ibrahim M.Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human.Dev Dyn.,2004,230:187–198.

[16] Zhang J, Espinoza LA and Kinders RJ. NANOG modulates stemness in human colorectal cancer.Oncogene. 2012 Oct 22.

[17] Zhou X, Zhou YP and Huang GR.Expression of the stem cell marker, Nanog, in human endometrial adenocarcinoma.Int J Gynecol Pathol,2011,30(3):262–270.

[18]Trapnell Cole,Pachter Lior and SalzbergL.Steven. TopHat: discovering splice junctions with RNA-Seq. Oxford JournalsLife Sciences & Mathematics & Physical Sciences Bioinformatics Volume 25, Issue 9Pp. 1105-1111.

[19] Trapnell C, Williams BA, Pertea G, Mortazavi AM, Kwan G, van Baren MJ, Salzberg SL, Wold B and Pachter L.Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms.Nature Biotechnology doi:10.1038/nbt.1621.

[20] Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J and Nekrutenko A; Galaxy Team. Manipulation of FASTQ data with Galaxy. Bioinformatics, 2010 Jul 15;26(14):1783-5.

[21] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009;4(1):44-57.

[22] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1-13.

[23] R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

[24] Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD and Ideker T.A travel guide to Cytoscape plugins.Nature Methods,2012 Nov;9(11):1069-76.

[25]Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T.Cytoscape: a software environment for integrated models of biomolecular interaction networks.Genome Research, 2003 Nov; 13(11):2498-504.